Neurosymbolic AI in CPS: Summary

Karthik Pattabiraman, UBC

Two Talks

Luis Garcia, Univ. of Utah

A Trip to the Neural Frontier: Neurosymbolic Sensor Fusion for Trustworthy CPS

Lu Feng, Univ. of Virginia

Predictive Monitoring and Safety Shielding for AI-Enabled CPS

Luis’s Talk

- Neuralink and brain computer interface (Elon Musk)
    - Cognitive state estimator + closed loop controller = AI-enabled Deep Brain Stimulation
    - What're the explainability and interpretability challenges of this stimulation?
    - Programmer: Safety guarantees, monitoring and feedback, patient-centered design

Challenges

- Traditional IoT: Low-dimensional structured sensor data. These have higher-dimensional data
    - Provide complex inferences from simple sensors. AI-enabled CPS
    - Resource constrained devices - need optimization to run complex DNNs on them
    - How should we explain DNNs? Post-hoc methods or interpretable DNNs
    - Most people preferred explanations that're post-hoc as long as examples were provided
        - The only exception was text data
- Concept based Interpretable DNNs (concepts bottleneck model)

Challenges (contd…)

Challenge 2: Combining data and knowledge (complex events)

- Bridging deep learning and symbolic models in AI-driven CPS (hybrid)
- Neuroplex inference: Deep learning perception + complex event reasoning
  - End-to-End training starting from complex events (e.g., washing hands)
  - Allows incorporating of human knowledge
  - Explainable complex human activity recognition
  - Needs humans to annotate the data with tasks

Recommended Reading: Neurosymbolic programming, Chaudhary

Back to the Neural Frontier

- Enhance human reasoning capabilities
- Both EEG readings & implanted sensors
- How humans encode memories (episodic)
- Helping people navigate complex buildings

Future challenges: Robustness, Security and Privacy. Explaining these to humans.

Lu Feng’s talk

Design time techniques not sufficient for safety guarantees.

- Need runtime techniques (predictive monitoring + safety shielding)
- Can predict the future state and monitor whether the future state will satisfy or violate the requirement
- Risk can be predicted for future states ("Predictive monitoring")
- Datasets for air quality monitoring from NYC
- Uncertainty in CPS arising from many factors (environment, human, noise)

Decision making under Uncertainty

- Decision making based on uncertain data. WHat should the decision maker do?

    - Relational RL models can capture uncertainty - model these as Gaussian distribution and find confidence intervals (95%)

    - Signal Temporal Logic with Uncertaining (STL-U): Can be applied to the air quality problem - 95% confidence, the predicted index < 100

    - When compared with no monitor, or STL-only monitoring, STL-U is able to do a much better job of preventing violations

Multi-agent Reinforcement Learning

- Multi-agent RL: Used in many CPS applications. Provide safety guarantees during safety guarantees

    - Safety-shielding for the multi-agent RL (MARL): Centralized Vs. Factored (run multiple shields in parallel after state space partitioning)

    - Specifications are expressed in LTL and shields are synthesized by solving two-player games

    - Evaluation results show this method is much better than centralized shielding. Also, centralized shielding doesn't work in continuous environments

Partially Observable Markov Decision Processes (POMDP)

Decision making under uncertainty. "Almost sure reach-avoid specification".

- Take an existing approach and integrate with POMDP. Also, factored shielding vs centralized shielding. The former has much higher scalability.

- If the obstacles are moving, then the problem becomes much more challenging. Prior work on "adaptive conformal prediction" partially addresses this.

- Safe AI-enabled CPS we need runtime techniques. Different AI methods need different safety guarantees.